Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment

Fonseca Guerra, C.; Bickelhaupt, F.M.; Snijders, J.G.; Baerends, E.J.

published in
Journal of the American Chemical Society
2000

DOI (link to publisher)
10.1021/ja993262d

document version
Publisher's PDF, also known as Version of record

Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment

Célia Fonseca Guerra,[†] F. Matthias Bickelhaupt,*[,†] Jaap G. Snijders,[‡] and Evert Jan Baerends[†]

Contribution from the Afdeling Theoretische Chemie, Scheikundig Laboratorium der Vrije Universiteit, De Boelelaan 1083, NL-1081 HV Amsterdam, The Netherlands, and Theoretische Chemie, Rijksuniversiteit Groningen, Nijenborgh 4, NL-9747 AG Groningen, The Netherlands

Received September 8, 1999. Revised Manuscript Received December 28, 1999

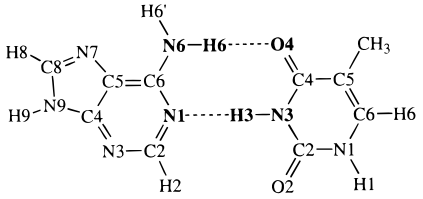
Abstract: Up till now, there has been a significant disagreement between theory and experiment regarding hydrogen bond lengths in Watson−Crick base pairs. To investigate the possible sources of this discrepancy, we have studied numerous model systems for adenine−thymine (AT) and guanine−cytosine (GC) base pairs at various levels (i.e., BP86, PW91, and BLYP) of nonlocal density functional theory (DFT) in combination with different Slater-type orbital (STO) basis sets. Best agreement with available gas-phase experimental A−T and G−C bond enthalpies (−12.1 and −21.0 kcal/mol) is obtained at the BP86/TZ2P level, which (for 298 K) yields −11.8 and −23.8 kcal/mol. However, the computed hydrogen bond lengths show again the notorious discrepancy with experimental values. The origin of this discrepancy is not the use of the plain nucleic bases as models for nucleotides: the disagreement with experiment remains no matter if we use hydrogen, methyl, deoxyribose, or 5′-deoxyribose monophosphate as the substituents at N9 and N1 of the purine and pyrimidine bases, respectively. Even the BP86/DZP geometry of the Watson−Crick-type dimer of deoxyadenylyl-3′,5′-deoxyuridine including one Na^+ ion (with 123 atoms our largest model for sodium adenylyl-3′,5′-uridine hexahydrate, the crystal of which had been studied experimentally with the use of X-ray diffraction) still shows this disagreement with experiment. The source of the divergence turns out to be the molecular environment (water, sugar hydroxyl groups, counterions) of the base pairs in the crystals studied experimentally. This has been missing, so far, in all theoretical models. After we had incorporated the major elements of this environment in our model systems, excellent agreement between our BP86/TZ2P geometries and the X-ray crystal structures was achieved.

1. Introduction

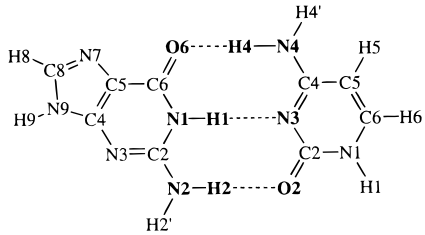
Hydrogen bonds are important in many fields of biological chemistry. They play, for instance, a key role in the working of the genetic code.[1] Already in 1953, Watson and Crick[1d] proposed a structure for DNA in which two helical chains of nucleotides are held together by the hydrogen bonds that occur in a selective fashion between a purine and a pyrimidine nucleic base giving rise to the Watson−Crick pairs adenine−thymine (AT) and guanine−cytosine (GC), see Scheme 1.

Until recently, hydrogen bonds were conceived as predominantly electrostatic phenomena that in the case of DNA base pairs are reinforced by polarization of the π-electron system (Resonance Assisted Hydrogen Bonding, RAHB).[2] Very recently,[3,4] we have shown through detailed analyses of the bonding mechanism that donor−acceptor orbital interactions between the DNA bases in the Watson−Crick pairs are of

Scheme 1



Adenine–Thymine



Guanine–Cytosine

* To whom correspondence should be addressed. Fax: +31−20−44 47 629. E-mail: bickel@chem.vu.nl.

[†] Scheikundig Laboratorium der Vrije Universiteit.

[‡] Rijksuniversiteit Groningen.

(1) (a) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, New York, Heidelberg, 1991. (b) Jeffrey, G. A. *An Introduction to Hydrogen bonding*; Oxford University Press: New York, Oxford, 1997, Chapter 10. (c) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, Berlin, Heidelberg, Tokyo, 1984. (d) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.

(2) (a) Gilli, P.; Ferretti, V.; Bertolasi, V.; Gilli, G., In: Advances in Molecular Structure Research; Hargittai, M., Hargittai, I., Eds.; JAI Press: Greenwich, CT, 1996; Vol. 2, pp 67−102. (b) Gilli, G.; Bellucci, F.; Ferretti, V.; Bertolasi, V. *J. Am. Chem. Soc.* **1989**, *111*, 1023. (c) Gilli, G.; Bertolasi, V.; Ferretti, V.; Gilli, P. *Acta Crystallogr.* **1993**, *B49*, 564.

(3) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem. Eur. J.* **1999**, *5*, 3581.

(4) Fonseca Guerra, C.; Bickelhaupt, F. M. *Angew. Chem.* **1999**, *111*, 3120; *Angew. Chem., Int. Ed.* **1999**, *38*, 2942.

10.1021/ja993262d CCC: $19.00
Published on Web 04/14/2000

comparable strength as electrostatic interactions. The donor−acceptor or charge-transfer term is provided by the interactions of lone-pair orbitals on O or N of one base with N−H σ^* acceptor orbitals of the other base. This picture complements and is in perfect agreement with experimental evidence[5] for a partial covalent character of hydrogen bonds obtained, lately, by different groups through X-ray diffraction studies on ice[5a] and NMR investigations on hydrogen bonds in RNA[5b] and in proteins.[5c,d]

In the present paper, we address a different point. Whereas both density functional and traditional ab initio methods satisfactorily reproduce experimental A−T and G−C hydrogen bond enthalpies,[6] there is a significant discrepancy between theory[7] and experiment[1c,8] regarding hydrogen bond lengths in the Watson−Crick base pairs (vide infra). Our purpose is to find and understand the source of this apparent disagreement using modern density functional theory (DFT) and, in this way, to arrive at a suitable quantum chemical approach for biochemical molecules that involve hydrogen bonds.

First, an extensive comparison is done between the performances of a number of density functionals (BP86, PW91, BLYP) in combination with different Slater-type orbital (STO) basis sets. The suitability of DFT for hydrogen-bonded systems has been the subject of many investigations.[9] In a study on the water dimer and the formamide−water complex, for example, Sim et al.[9a] found that nonlocal DFT performs satisfactorily, yielding results that compare well with those from correlated ab initio methods. At this point, however, we anticipate that whereas our highest-level base-pairing enthalpies are in excellent agreement with gas-phase experimental values, we still arrive at the notorious discrepancies with experimental (X-ray crystal) structures that were encountered before.

In a preliminary communication,[4] we have briefly reported how this, and the fact that Watson−Crick base pairing is associated with very shallow potential energy surfaces, has led us to study the possible effects of using better models for the glycosidic N−C bond as well as the influence of the molecular environment that the bases experience in the crystals studied experimentally.[1c,8] Here, we present a full account of our investigations which have been extended, meanwhile, to larger DNA segments. In particular, we present the results of the first high-level DFT study of the Watson−Crick-type dimer of adenylyl-3′,5′-uridine, i.e., $(ApU)_2$, including a full geometry optimization and bond analysis. Furthermore, the structure of and bonding in Watson−Crick pairs of, among others, methylated bases, nucleosides and nucleotides are examined.

It is important to note that the experimental AT (or AU) and GC structures were obtained from X-ray diffraction studies on crystals of sodium adenylyl-3′,5′-uridine hexahydrate and sodium guanylyl-3′,5′-cytidine nonahydrate.[1c,8] In these crystals, the functional groups of the DNA bases that are involved in Watson−Crick hydrogen bonding can also enter (hydrogen bonding) interactions with water molecules, hydroxyl groups of sugar residues, and Na^+ counterions. Recent theoretical studies have furthermore demonstrated that the structure of Watson−Crick pairs can be significantly influenced by the interaction with a metal cation.[7k−q] Thus, we simulate the crystal environment by incorporating up to six water molecules (modeling both crystal water and sugar OH groups) and one Na^+ ion into our model systems. Also the effect was studied of a sodium ion on the hydrogen bonds in the adenylyl-3′,5′-uridine pair, $(ApU)_2$. We are interested in both the structural and energetic consequences, and we try to rationalize them in terms of the Kohn−Sham MO model and by analyzing the electron density redistribution associated with particular chemical interactions.

2. Theoretical Methods

2A. General Procedure. All calculations were performed using the Amsterdam Density Functional (ADF) program[10] developed by Baerends et al.[10a−d] and parallelized[10a] as well as linearized[10e] by Fonseca Guerra et al. The numerical integration was performed using the procedure developed by Boerrigter, te Velde, and Baerends.[10f,g] The MOs were expanded in a large uncontracted set of Slater type orbitals (STOs) containing diffuse functions: DZP and TZ2P.[10h] The DZP basis set is of double-ζ quality for all atoms and has been augmented with one set of polarization functions: 3d on C, N, O; and 2p on H. The TZ2P basis set is of triple-ζ quality for all atoms and has been augmented with two sets of polarization functions: 3d and 4f on C, N, O, P; and 2p and 3d on H. The 1s core shell of carbon, nitrogen, oxygen and the 1s 2s 2p core shells of phosphorus were treated by the frozen-core (FC) approximation.[10b] An auxiliary set of s, p, d, f, and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately in each SCF cycle.[10i]

Energies and geometries were calculated at four different levels of theory: (i) the local density approximation (LDA), where exchange is described by Slater's Xα potential[10j] with $\alpha = {}^2/_3$ and correlation is treated in the Vosko−Wilk−Nusair (VWN) parametrization;[10k] (ii) LDA with nonlocal corrections to exchange due to Becke[10l,m] and correlation

(5) (a) Isaacs, E. D.; Shukla, A.; Platzman, P. M.; Hamann, D. R.; Barbiellini, B.; Tulk, C. A. *Phys. Rev. Lett.* **1999**, *82*, 600. (b) Dingley, A. J.; Grzesiek, S. *J. Am. Chem. Soc.* **1998**, *120*, 8293. (c) Cordier, F.; Grzesiek, S. *J. Am. Chem. Soc.* **1999**, *121*, 1601. (d) Cornilescu, G.; Hu, J.-S.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 2949.

(6) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers* **1979**, *18*, 1149.

(7) (a) Santamaria, R.; Vázquez, A. *J. Comput. Chem.* **1994**, *15*, 981. (b) Bertran, J.; Oliva, A.; Rodríguez-Santiago, L.; Sodupe, M. *J. Am. Chem. Soc.* **1998**, *120*, 8159. (c) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493. (d) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965. (e) Brameld, K.; Dasgupta, S.; Goddard, W. A., III. *J. Phys. Chem. B* **1997**, *101*, 4851. (f) Hobza, P.; Sponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379. For related topics, see also: (g) Sponer, J.; Hobza, P. In *Encyclopedia of Computational Chemistry*; von Ragué Schleyer, P., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley and Sons: Chichester, New York, Weinheim, Brisbane, Singapore, Toronto, 1998; Vol 1, pp 777−789. (h) Sponer, J.; Hobza, P.; Leszczynski, J. In *Computational Chemistry. Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific Publisher: Singapore, 1996; pp 185−218. (i) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590. (j) Alhambra, C.; Luque, F. J.; Gago, F.; Orozco, M. *J. Phys. Chem. B* **1997**, *101*, 3846. (k) Sponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1999**, *103*, 2528. (l) Sponer, J.; Burda, J. V.; Sabat, M.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 5951. (m) Burda, J. V.; Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670. (n) Anwander, E. H. S.; Probst, M. M.; Rode, B. M. *Biopolymers* **1990**, *29*, 757. (o) Sagarik, K. P.; Rode, B. M. *Inorg. Chim. Acta* **1983**, *78*, 177. (p) Burda, J. V.; Sponer, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 7250. (q) Basch, H.; Krauss, M.; Stevens, W. J. *J. Am. Chem. Soc.* **1985**, *107*, 7267.

(8) (a) Seeman, N. C.; Rosenberg, J. M.; Suddath, F. L.; Kim, J. J. P.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 109. (b) Rosenberg, J. M.; Seeman, N. C.; Day, R. O.; Rich, A. *J. Mol. Biol.* **1976**, *104*, 145.

(9) (a) Sim, F.; St-Amant, A.; Papai, I.; Salahub, D. R. *J. Am. Chem. Soc.* **1992**, *114*, 4391. (b) Guo, H.; Sirois, S.; Proynov, E. I.; Salahub, D. R. In *Theoretical Treatment of Hydrogen Bonding*; Hadzi, D., Ed.; Wiley: New York, 1997. (c) Sirois, S.; Proynov, E. I.; Nguyen, D. T.; Salahub, D. R. *J. Chem. Phys.* **1997**, *107*, 6770. (d) Kim, K.; Jordan, K. D. *J. Phys. Chem.* **1994**, *98*, 10089. (e) Novoa, J. J.; Sosa, C. *J. Phys. Chem.* **1995**, *99*, 15837. (f) Latajka, Z.; Bouteiller, Y. *J. Chem. Phys.* **1994**, *101*, 9793. (g) Del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705. (h) Florian, J.; Johnson, B. G. *J. Phys. Chem.* **1995**, *99*, 5899. (i) Combariza, J. E.; Kestner, N. R. *J. Phys. Chem.* **1995**, *99*, 2717. (j) Civalleri, B.; Garrone, E.; Ugliengo, P. *J. Mol. Struct.* **1997**, *419*, 227. (k) Lozynski, M.; Rusinska-Roszak, D.; Mack, H.-G. *J. Phys. Chem.* **1998**, *102*, 2899. (l) Chandra, A. K.; Nguyen, M. *Chem. Phys.* **1998**, *232*, 299. (m) Paizs, B.; Suhai, S. *J. Comput. Chem.* **1998**, *19*, 575. (n) McAllister, M. A. *J. Mol. Struct.* **1998**, *427*, 39. (o) Pan, Y. P.; McAllister, M. A. *J. Mol. Struct.* **1998**, *427*, 221. (p) Gonzalez, L.; Mo, O.; Yanez, M. *J. Comput. Chem.* **1997**, *18*, 1124. (q) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem.* **1998**, *102*, 3782.

due to Perdew[10n] added self-consistently[10o] (BP86); (iii) LDA with nonlocal corrections to exchange and correlation due to Perdew and Wang[10p,q] also added self-consistently (PW91); (iv) LDA with nonlocal corrections to exchange due to Becke[10m] and correlation due to Lee−Yang−Parr[10r,s] added, again, self-consistently (BLYP).

Geometries were optimized using analytical gradient techniques.[10t] Frequencies[10u] were calculated by numerical differentiation of the analytical energy gradients and using the nonlocal density functionals. The basis set superposition error (BSSE), associated with the hydrogen bond energy, has been computed via the counterpoise method,[10v] using the individual bases as fragments. Bond enthalpies at 298.15 K and 1 atm (ΔH_{298}) were calculated from 0 K electronic bond energies (ΔE) according to eq 1, assuming an ideal gas.[11]

$$\Delta H_{298} = \Delta E + \Delta E_{\mathrm{trans},298} + \Delta E_{\mathrm{rot},298} + \Delta E_{\mathrm{vib},0} + \Delta(\Delta E_{\mathrm{vib}})_{298} + \Delta(pV) \quad (1)$$

Here, $\Delta E_{\mathrm{trans},298}$, $\Delta E_{\mathrm{rot},298}$, and $\Delta E_{\mathrm{vib},0}$ are the differences between products and reactants in translational, rotational, and zero point vibrational energy, respectively; $\Delta(\Delta E_{\mathrm{vib}})_{298}$ is the change in the vibrational energy difference as one goes from 0 to 298.15 K. The vibrational energy corrections are based on our frequency calculations. The molar work term $\Delta(pV)$ is $(\Delta n)RT$; $\Delta n = -1$ for two fragments combining to one molecule. Thermal corrections for the electronic energy are neglected.

2B. Bond Analysis. The bonding in the various AT and GC model systems was analyzed in the conceptual framework provided by the Kohn−Sham molecular orbital (KS-MO) model[12] using the extended transition state (ETS) method developed by Ziegler and Rauk to decompose the bond energy.[13] The overall bond energy ΔE is made up of two major components (eq 2).

$$\Delta E = \Delta E_{\mathrm{prep}} + \Delta E_{\mathrm{int}} \quad (2)$$

The preparation energy ΔE_{prep} is the amount of energy required to deform the separate molecular fragments (e.g., nucleic bases) from their equilibrium structure to the geometry they acquire in the composite system (e.g., the base pair). The interaction energy ΔE_{int} corresponds to the actual energy change when the prepared fragments are combined to form the composite system. It is further split up into three physically meaningful terms (eq 3):

$$\Delta E_{\mathrm{int}} = \Delta V_{\mathrm{elstat}} + \Delta E_{\mathrm{Pauli}} + \Delta E_{\mathrm{oi}} \quad (3)$$

The term $\Delta V_{\mathrm{elstat}}$ corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the prepared fragments and is usually attractive. The Pauli-repulsion $\Delta E_{\mathrm{Pauli}}$ comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsion. The orbital interaction ΔE_{oi} accounts for charge transfer (interaction between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO−LUMO interactions) and polarization (empty/occupied orbital mixing on one fragment due to the presence of another fragment). It can be decomposed into the contributions from each irreducible representation Γ of the interacting system (eq 4).[13] In systems with a clear σ, π separation (e.g., flat DNA base pairs) this symmetry partitioning proves to be most informative.

$$\Delta E_{\mathrm{oi}} = \sum_{\Gamma} \Delta E_{\Gamma} \quad (4)$$

2C. Analysis of the Charge Distribution. The electron density distribution is analyzed using the Voronoi deformation density (VDD) method introduced in ref 14 and further developed in ref 3 to enable a correct treatment of even the subtle changes in atomic charges $\Delta Q_A^{\mathrm{VDD}}$ caused by weak chemical interactions (such as hydrogen bonds) between molecular fragments as well as a decomposition into the contributions from the σ- and π-electron systems. VDD atomic charges $\Delta Q_A^{\mathrm{VDD}}$ are defined and related to the deformation density $\rho_{\mathrm{total\ system}}(\mathbf{r}) - \rho_{\mathrm{subsystem1}}(\mathbf{r}) - \rho_{\mathrm{subsystem2}}(\mathbf{r})$ by eq 5.[3]

$$\Delta Q_A^{\mathrm{VDD}} = -\int_{\substack{\text{Voronoi cell of A} \\ \text{in total system}}} (\rho_{\mathrm{total\ system}}(\mathbf{r}) - \rho_{\mathrm{subsystem1}}(\mathbf{r}) - \rho_{\mathrm{subsystem2}}(\mathbf{r}))d\mathbf{r} \quad (5)$$

The interpretation of VDD atomic charges is rather straightforward. Instead of measuring the amount of charge associated with a particular atom A, they directly monitor how much charge flows, due to chemical interactions, out of ($\Delta Q_A^{\mathrm{VDD}} > 0$) or into ($\Delta Q_A^{\mathrm{VDD}} < 0$) the Voronoi cell of atom A, that is, the region of space that is closer to nucleus A than to any other nucleus. The Voronoi cell of atom A is bounded by the bond midplanes on and perpendicular to all bond axes between nucleus A and its neighboring nuclei (cf., the Wigner−Seitz cells in crystals).[10g,15]

3. Watson−Crick Pairs of Plain Nucleic Bases

3A. Hydrogen Bond Strength. To examine the performance of the different density functionals and STO basis sets, we have studied the formation of the plain adenine−thymine and guanine−cytosine complexes (see Scheme 1) at five different levels of theory: BP86/TZ2P, BLYP/TZ2P, PW91/TZ2P, BP86/DZP, and also LDA/TZ2P. In Tables 1−4, the results of our calculations are summarized and compared with those from other theoretical[7] and experimental[1c,6,8] studies. The LDA functional leads to overbinding (i.e., hydrogen bonds are too short and too strong at LDA/TZ2P; data not shown here), in line with general experience, and will not be further discussed.

Tables 1 and 2 show that BP86/TZ2P provides A−T and G−C bond enthalpies (−11.8 and −23.8 kcal/mol) that agree within 0.3 and 2.8 kcal/mol, respectively, with the corresponding gas-phase experimental[6] values (−12.1 and −21.0 kcal/mol). The PW91/TZ2P and the BLYP/TZ2P levels yield bond enthalpies that deviate somewhat more from experiment: they are too strongly binding by 1−2 kcal/mol for AT and by some

(10) (a) Fonseca Guerra, C.; Visser, O.; Snijders, J. G.; te Velde, G.; Baerends, E. J. In *Methods and Techniques for Computational Chemistry*; Clementi, E., Corongiu, G., Eds.; STEF: Cagliari, 1995; pp 305−395. (b) Baerends, E. J.; Ellis, D. E.; Ros, P. *Chem. Phys.* **1973**, *2*, 41. (c) Baerends, E. J.; Ros, P. *Chem. Phys.* **1975**, *8*, 412. (d) Baerends, E. J.; Ros, P. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1978**, *S12*, 169. (e) Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Acc.* **1998**, *99*, 391. (f) Boerrigter, P. M.; te Velde, G.; Baerends, E. J. *Int. J. Quantum Chem.* **1988**, *33*, 87. (g) te Velde, G.; Baerends, E. J. *J. Comput. Phys.* **1992**, *99*, 84. (h) Snijders, J. G.; Baerends, E. J.; Vernooijs, P. *At. Nucl. Data Tables* **1982**, *26*, 483. (i) Krijn, J.; Baerends, E. J. *Fit-Functions in the HFS-Method; Internal Report* (in Dutch); Vrije Universiteit: Amsterdam, 1984. (j) Slater, J. C. *Quantum Theory of Molecules and Solids Vol. 4*; McGraw-Hill: New York, 1974. (k) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200. (l) Becke, A. D. *J. Chem. Phys.* **1986**, *84*, 4524. (m) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098. (n) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822 (Erratum: *Phys. Rev. B* **1986**, *34*, 7406). (o) Fan, L.; Ziegler, T. *J. Chem. Phys.* **1991**, *94*, 6057. (p) Perdew, J. P. In *Electronic Structure of Solids*; Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; pp 11−20. (q) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1992**, *46*, 6671. (r) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785. (s) Johnson, B. G.; Gill, P. M. W.; Pople, J. A. *J. Chem. Phys.* **1992**, *98*, 5612. (t) Versluis, L.; Ziegler, T. *J. Chem. Phys.* **1988**, *88*, 322. (u) Fan, L.; Versluis, L.; Ziegler, T.; Baerends, E. J.; Ravenek, W. *Int. J. Quantum. Chem., Quantum. Chem. Symp.* **1988**, *S22*, 173. (v) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(11) Atkins, P. W. *Physical Chemistry*; Oxford University Press: Oxford, 1982.

(12) Bickelhaupt, F. M.; Baerends, E. J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2000; Vol. 15, Chapter 1.

(13) (a) Bickelhaupt, F. M.; Nibbering, N. M. M.; van Wezenbeek, E. M.; Baerends, E. J. *J. Phys. Chem.* **1992**, *96*, 4864. (b) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1755. (c) Ziegler, T.; Rauk, A. *Inorg. Chem.* **1979**, *18*, 1558. (d) Ziegler, T.; Rauk, A. *Theor. Chim. Acta* **1977**, *46*, 1.

(14) Bickelhaupt, F. M.; van Eikema Hommes, N. J. R.; Fonseca Guerra, C.; Baerends, E. J. *Organometallics* **1996**, *15*, 2923.

(15) Kittel, C. *Introduction to Solid State Physics*; Wiley: New York, 1986.

Table 1. Hydrogen Bond Strength (kcal/mol) in AT (plain nucleic bases, unless stated otherwise)

	ΔE^{AT}	$\Delta E^{AT}_{(BSSE)}$	ΔE^{AT}_{298}
experiment[a]			−12.1
DFT with STO basis (this work)			
BP86/TZ2P (C_1)[b]	−13.0	−12.3	−11.8
BP86/TZ2P (pair: C_s, bases: C_1)[c]	−13.0	−12.3	
BP86/TZ2P (C_s)[d]	−13.0	−12.3	
BP86/DZP (C_1)[b]	−15.8	−12.7	
PW91/TZ2P (C_1)[b]	−15.2	−14.5	−14.0[j]
BLYP/TZ2P (C_1)[b]	−14.5	−13.7	−13.2[j]
DFT with GTO basis (others)			
BP86/DZVP (C_1)[b,e]		−13.9	
B3LYP/6-31G** (C_1)[b,f]		−12.3	−10.9
ab initio with GTO basis (others)			
MP2/DZP//HF/6-31G* (C_1)[b,g]		−14.7	−11.9
MP2/6-31G*(0.25)//HF/6-31G** (C_s)[d,h]		−11.8	−9.5
LMP2/cc-pVTZ(-f)//HF/cc-pVTZ(-f) (C_1)[i]		−10.8	−10.2

[a] ΔH_{exp} from mass spectrometry data of Yanson et al.[6] for AT with 9-methyladenine and 1-methylthymine (**1c**) with corrections according to Brameld et al.[7e] [b] Full optimization in C_1 symmetry of base pair and bases. [c] Base pair optimized in C_s (**1a**) and bases in C_1 symmetry. [d] Optimization in C_s symmetry of base pair (**1a**) and bases. [e] Santamaria et al.[7a] [f] Bertran et al.[7b] [g] Gould et al.;[7c] AT with 9-methyladenine and 1-methylthymine (**1c**). [h] Sponer et al.[7d,f] [i] Brameld et al.;[7e] AT with 9-(hydroxymethyl)adenine and 1-(hydroxymethyl)thymine. Base pair optimized in C_1 and bases in C_s symmetry. [j] ΔH_{298}^{AT} was obtained with thermal energy corrections from BP86/TZ2P (C_1).

Table 2. Hydrogen Bond Strength (kcal/mol) in GC (plain nucleic bases, unless stated otherwise)

	ΔE^{GC}	$\Delta E^{GC}_{(BSSE)}$	ΔH^{GC}_{298}
experiment[a]			−21.0
DFT with STO basis (this work)			
BP86/TZ2P (C_1)[b]	−26.1	−25.2	−23.8
BP86/TZ2P (pair: C_s, bases: C_1)[c]	−26.1	−25.2	
BP86/TZ2P (C_s)[d]	−26.5	−25.7	
BP86/DZP (C_1)[b]	−28.8	−25.1	
PW91/TZ2P (C_1)[b]	−28.5	−27.7	−26.3[j]
BLYP/TZ2P (C_1)[b]	−28.3	−27.4	−26.0[j]
DFT with GTO basis (others)			
BP86/DZVP (C_1)[b,e]		−27.7	
B3LYP/6-31G** (C_1)[b,f]		−25.5	−24.0
ab initio with GTO basis (others)			
MP2/DZP//HF/6-31G* (C_1)[b,g]		−28.0	−25.4
MP2/6-31G*(0.25)//HF/6-31G** (C_s)[d,h]		−23.4	−20.8
LMP2/cc-pVTZ(-f)//HF/cc-pVTZ(-f) (C_1)[i]		−22.4	−21.2

[a] ΔH_{exp} from mass spectrometry data of Yanson et al.[6] for GC with 9-methylguanine and 1-methylcytosine (**2b**). [b] Full optimization in C_1 symmetry of base pair and bases. [c] Base pair optimized in C_s (**2a**) and bases in C_1 symmetry. [d] Optimization in C_s symmetry of base pair (**2a**) and bases. [e] Santamaria et al.[7a] [f] Bertran et al.[7b] [g] Gould et al.;[7c] GC with 9-methylguanine and 1-methylcytosine (**2b**). [h] Sponer et al.[7d,f] [i] Brameld et al.;[7e] GC with 9-(hydroxymethyl)guanine and 1-(hydroxymethyl)cytosine. Base pair optimized in C_1 and bases in C_s symmetry. [j] ΔH_{298}^{GC} was obtained with thermal energy corrections from BP86/TZ2P (C_1).

5 kcal/mol for GC. Also the DFT and ab initio results of others agree reasonably well with experiment (Tables 1 and 2).

Whereas the basis set superposition error (BSSE) is less than 1 kcal/mol in the case of the TZ2P basis, it rises to 3.7 kcal/mol if the smaller DZP basis is used [compare BP86/TZ2P (C_1) and BP86/DZP (C_1) entries in Tables 1 and 2]. Note, however, that the BSSE-corrected BP86/DZP values agree within 0.4 (AT) or 0.1 kcal/mol (GC). This suggests that we can use the DZP basis as a suitable and efficient alternative to the much larger TZ2P basis for studying very large systems involving hydrogen bonds, provided that energies are corrected for the BSSE.

We have also studied the effect of symmetry constraints on the base-pairing energies by examining, at BP86/TZ2P, three

Table 3. Hydrogen Bond Distances (Å) in AT (plain nucleic bases, unless stated otherwise)

	N6−O4	N1−N3
experiment[a]		
A2U1	2.95	2.82
A1U2	2.93	2.85
DFT with STO basis (this work)		
BP86/TZ2P (C_1)[b]	2.85	2.81
BP86/TZ2P (C_s)[c]	2.85	2.81
BP86/DZP (C_1)[b]	2.84	2.79
PW91/TZ2P (C_1)[b]	2.85	2.79
BLYP/TZ2P (C_1)[b]	2.84	2.78
DFT with GTO basis (others)		
BP86/DZVP (C_1)[b,d]	2.95	2.87
B3LYP/6-31G** (C_1)[b,e]	2.94	2.84
ab initio with GTO basis (others)		
HF/6-31G* (C_1)[b,f]	3.08	3.01
HF/6-31G** (C_s)[c,g]	3.09	2.99
HF/cc-pVTZ(-f) (C_1)[b,h]	3.06	2.92

[a] X-ray crystallographic measurements by Seeman et al.[8a] on sodium adenylyl-3′,5′-uridine hexahydrate (**1**) containing the Watson−Crick-type dimer $(ApU)_2$. There are two values for each hydrogen bond length because the two AU pairs (A2U1 and A1U2) have different environments (see Scheme 2). [b] Full optimization in C_1 symmetry. [c] Optmization in C_s symmetry (**1a**). [d] Santamaria et al.[7a] [e] Bertran et al.[7b] [f] Gould et al.;[7c] AT with 9-methyladenine and 1-methylthymine (**1c**). [g] Sponer et al.[7d] [h] Brameld et al.;[7e] AT with 9-(hydroxymethyl)adenine and 1-(hydroxymethyl)thymine.

Table 4. Hydrogen Bond Distances (Å) in GC (plain nucleic bases, unless stated otherwise)

	O6−N4	N1−N3	N2−O2
experiment			
GpC [a]	2.91	2.95	2.86
DFT with STO basis (this work)			
BP86/TZ2P (C_1)[b]	2.73	2.88	2.87
BP86/TZ2P (C_s)[c]	2.73	2.88	2.87
BP86/DZP (C_1)[b]	2.71	2.87	2.87
PW91/TZ2P (C_1)[b]	2.72	2.88	2.87
BLYP/TZ2P (C_1)[b]	2.71	2.86	2.84
DFT with GTO basis (others)			
BP86/DZVP (C_1)[b,d]	2.78	2.93	2.93
B3LYP/6-31G** (C_1)[b,e]	2.79	2.93	2.92
ab initio with GTO basis (others)			
HF/6-31G* (C_1)[b,f]	2.93	3.05	3.01
HF/6-31G** (C_s)[c,g]	2.92	3.04	3.02
HF/cc-pVTZ(-f) (C_1)[b,h]	2.83	2.95	2.92

[a] X-ray crystallographic measurements by Rosenberg et al.[8b] on sodium guanylyl-3′,5′-cytidine nonahydrate (**2**) containing the Watson−Crick-type dimer $(GpC)_2$. [b] Full optimization in C_1 symmetry. [c] Optimization in C_s symmetry (**2a**). [d] Santamaria et al.[7a] [e] Bertran et al.[7b] [f] Gould et al.;[7c] GC with 9-methylguanine and 1-methylcytosine (**2b**). [g] Sponer et al.[7d] [h] Brameld et al.;[7e] GC with 9-(hydroxymethyl)guanine and 1-(hydroxymethyl)cytosine.

different situations (see Tables 1 and 2): (i) both the base pair and the separate bases are fully optimized in C_1 symmetry; (ii) a C_s symmetry constraint applies to the base pair, but the bases are still fully optimized in C_1 symmetry; and (iii) a C_s symmetry restriction applies to both base pair and bases. There appears to be virtually no difference between base-pairing energies computed according to (i) and (ii) but those of (iii) deviate slightly (by half a kcal/mol) in the case of GC. This is not difficult to understand: whereas the Watson−Crick base pairs and also, although to a slightly lesser extent, the bases adenine, thymine, and cytosine are nearly planar, in guanine the pyramidalization of the N2 amino group is quite pronounced (see section 3B). Thus, reliable bond energies can be obtained efficiently from C_s-optimized Watson−Crick pairs (provided the separate bases are fully optimized). These results also show that the A−T and G−C bond analyses can be carried out in C_s

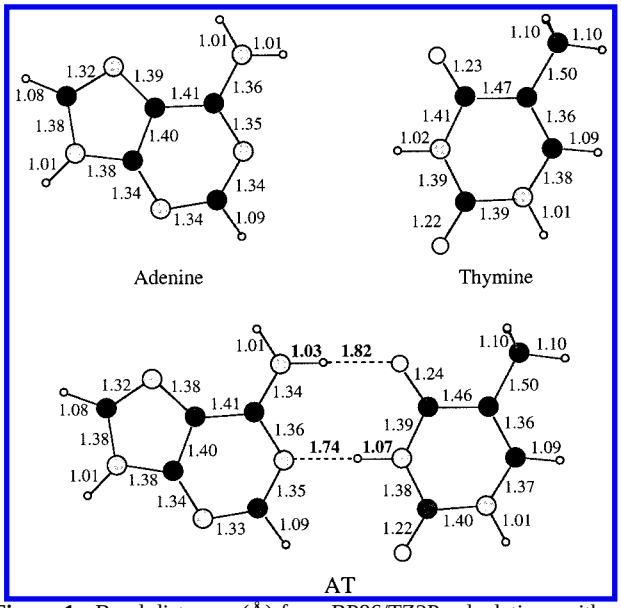


Figure 1. Bond distances (Å) from BP86/TZ2P calculations without any symmetry constraint for adenine, thymine, and the Watson−Crick pair AT (see Scheme 1).

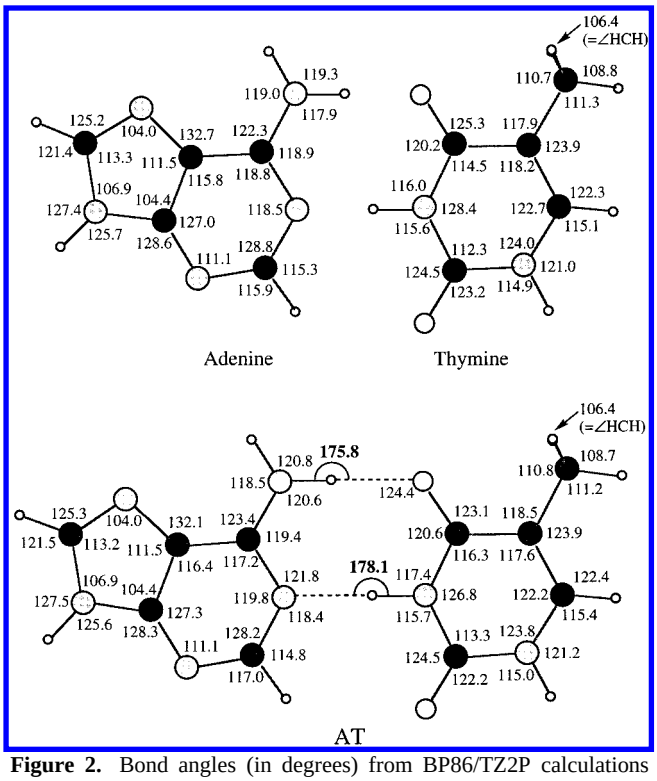


Figure 2. Bond angles (in degrees) from BP86/TZ2P calculations without symmetry constraints for adenine, thymine, and the Watson−Crick pair AT (see Scheme 1).

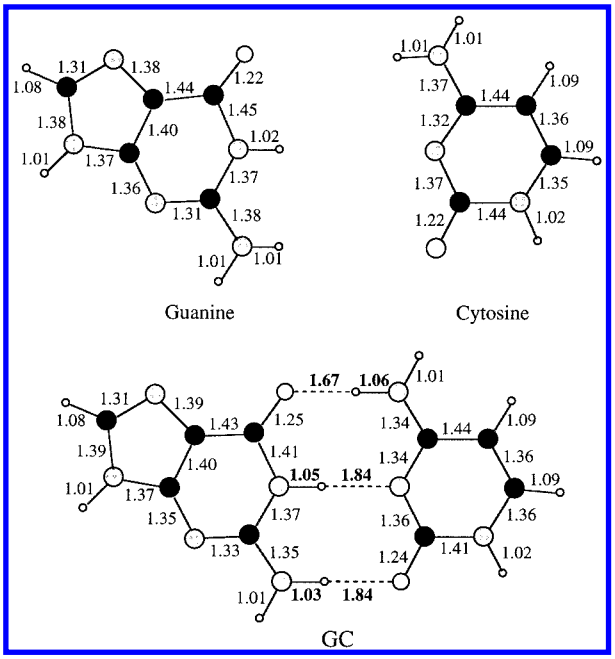


Figure 3. Bond distances (Å) from BP86/TZ2P calculations without any symmetry constraint for guanine, cytosine, and the Watson−Crick pair GC (see Scheme 1).

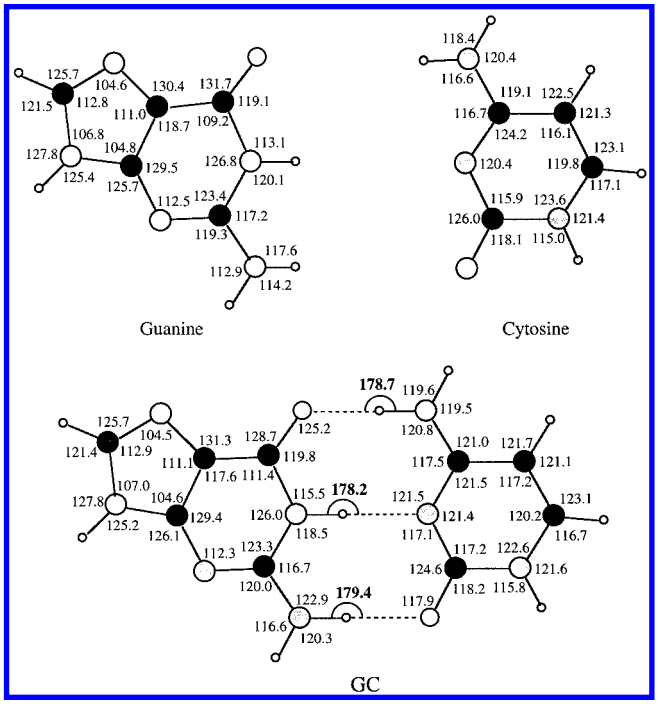


Figure 4. Bond angles (in degrees) from BP86/TZ2P calculations without symmetry constraints for guanine, cytosine, and the Watson−Crick pair GC (see Scheme 1).

symmetry, which allows for a decomposition of orbital interactions into σ and π contributions (eq 4).

3B. Structure of Bases and Watson−Crick Pairs. The BP86/TZ2P geometries of Watson−Crick pairs and separate bases are shown in Figures 1−4. All structures have been verified to be energy minima through vibrational analyses (no imaginary frequencies). As mentioned above, the amino groups of adenine (∠H6′N6C6C5 and ∠H6N6C6N1 are 11.9° and 11.4°) and cytosine (∠H4′N4C4C5 and ∠H4N4C4N3 are 15.1° and 10.6°) are only slightly pyramidal whereas that of guanine (∠H2′N2C2N3 and ∠H2N2C2N1 are 13.2° and 33.1°!) is more strongly pyramidalized in line with previous results[16] (we give absolute values of dihedral angles). The base pairs, in which the amino groups adopt a planar conformation, deviate only slightly from C_s symmetry. Furthermore, the hydrogen bonds in AT (∠N6H6O4 and ∠N1H3N3 are 175.8° and 178.1°) and GC (∠O6H4N4, ∠N1H1N3 and ∠N2H2O2 are 178.7°, 178.2° and 179.4°) are essentially linear (see Figures 2 and 4).

(16) (a) Sponer, J.; Hobza, P.; Leszczynski, J. In *Computational Chemistry. Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific Publisher: Singapore, 1996; pp 185−218. (b) Sponer, J.; Hobza, P. *Int. J. Quantum Chem.* **1996**, *57*, 959. (c) Stewart, E. L.; Foley, C. K.; Allinger, N. L.; Bowen, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 7282. (d) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161.

Regarding the Watson−Crick hydrogen bond distances, we arrive at striking discrepancies with experimental structures (see Tables 3 and 4). At BP86/TZ2P, we find N6−O4 and N1−N3 hydrogen bond distances in AT of 2.85 and 2.81 Å; this result is not influenced by applying a C_s symmetry constraint (Table 3). Only slightly shorter N6−O4 and N1−N3 distances are obtained with the smaller DZP basis (i.e., BP86/DZP) and with the other nonlocal functional (PW91/TZ2P and BLYP/TZ2P). These values have to be compared with 2.95 and 2.82 Å from experiment.[8a] Even more eye-catching, as can be seen in Table 4, is the situation for the three hydrogen bonds in GC, i.e., O6−N4, N1−N3, and N2−O2, for which we find a bond length pattern that is short−long−long, i.e., 2.73, 2.88, and 2.87 Å at BP86/TZ2P, at significant variance with the experimental values[8b] which are long−long−short (2.91, 2.95, and 2.86 Å). Again, the smaller DZP basis (i.e., BP86/DZP) and the other nonlocal functionals (PW91/TZ2P and BLYP/TZ2P) perform comparably well, yielding hydrogen bonds that are only slightly (i.e., 0.01−0.03 Å) shorter. Strikingly, our BP86/TZ2P approach does perform highly satisfactorily in case of the water dimer, $H_2O{\cdot}HOH$, if compared with the various high-level ab initio benchmark computations[17] available for this archetype hydrogen-bonded complex. Our BSSE-corrected hydrogen bond energy of −4.3 kcal/mol agrees, for example, within a few tenths of a kcal/mol with values obtained at MP2/aug-cc-pVQZ (−4.9 kcal/mol)[17a] and CCSD(T)/daug-cc-pVQZ (−5.1 kcal/mol; with counterpoise correction: −4.9 kcal/mol).[17b] More importantly, our O−O distance of 2.90 Å matches those from MP2/aug-cc-pVQZ (2.90 Å; with counterpoise correction: 2.92 Å)[17a] and CCSD(T) computations (2.90 Å;[17b] with counterpoise correction: [17c] 2.93 Å) either exactly or within three hundreths of an angstrom.

The disagreement between theoretical and experimental[1c,8] Watson−Crick hydrogen bond length is not new. It has been encountered before in several DFT and ab initio studies (see Tables 3 and 4).[7] For example, Hartree−Fock (HF)[7c−e] gives hydrogen bonds that are up to 0.2 Å longer than both our computed and the experimental[8] values, and in the case of GC the wrong bond length pattern of long−long−short is found (Table 4). Furthermore, whereas the DFT results of Bertran et al.[7b] and Santamaria et al.[7a] for AT are in good agreement with experimental structures,[8a] their geometries for GC differ again significantly from experimental ones.[8b] Differences between our and the other DFT geometries (the latter are up to 0.1 Å longer) can be ascribed, among others, to the use of different basis sets: STOs in our calculations and GTOs in those of Bertran et al.[7b] and Santamaria et al.[7a]

It is important to realize that the experimental structures stem from X-ray diffraction measurements on crystals of sodium adenylyl-3′,5′-uridine hexahydrate (**1**)[8a] for AT (or AU) and sodium guanylyl-3′,5′-cytidine nonahydrate (**2**)[8b] for GC. The base pairs in these crystals differ from the theoretical model systems studied so far, in two important fashions: (i) they are part of a small double helix consisting of two base pairs in which bases along a strand are connected via a ribose−phosphate−ribose backbone, and (ii) they experience interactions with the environment in the crystal, in particular water molecules, ribose OH groups, and counterions. In view of the very shallow potential energy surfaces that we find for Watson−Crick base pairing, it seems plausible that the effects of the backbone and the molecular environment in the crystal could cause the discrepancy with more simplistic AT and GC models. This has led us to study the effect of the backbone (section 4) and the molecular environment (section 5) at the BP86/TZ2P level which yields our best hydrogen bond enthalpies (see section 3A).

(17) For high-level ab initio benchmark computations on the water dimer, see, for example: (a) Hobza, P.; Bludsky, O.; Suhai, S. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3073. (b) Halkier, A.; Koch, H.; Jørgensen, P.; Christiansen, O.; Nielsen, I. M. B.; Helhaker, T. *Theor. Chem. Acc.* **1997**, *97*, 150. (c) Schütz, M.; Brdarski, S.; Widmark, P.-O.; Lindh, R.; Karlström, G. *J. Chem. Phys.* **1997**, *107*, 4597. (d) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821. (e) van Duineveldt-van de Rijdt, J. G. C. M.; van Duineveldt, F. B. *J. Chem. Phys.* **1999**, *111*, 3812. (f) Klopper, W.; Lüthi, H. P. *Mol. Phys.* **1999**, *96*, 559. For experimental studies, see: (g) Odutola, J. A.; Dyke, T. R. *J. Chem. Phys.* **1980**, *72*, 5062. (h) Curtiss, L. A.; Frurip, D. J.; Blander, M. *J. Chem. Phys.* **1979**, *71*, 2703. Note that the experimental (microwave) O−O distance of 2.95−2.98 Å (ref 17g) has been questioned on the basis of the accurate ab initio studies (see, e.g., refs 17a−c), which suggest an O−O distance that is up to 0.05 Å shorter than the lower bound. The experimental value is probably too long due to the underestimation of anharmonic corrections.

4. The Effect of the Backbone

The effect of the backbone on Watson−Crick base pairing is studied stepwise by going from Watson−Crick pairs of plain nucleic bases via nucleotides to strands consisting of two nucleotides (**1a**−**e**, **1k**, **2a**−**d**). The results are summarized in Figures 5−7 (geometries) and Tables 5 and 6 (Watson−Crick-pairing energies). In the first place, comparison of plain AT (**1a**) and AU (**1b**) shows that methylation at C5 of uracil has basically no influence on Watson−Crick pairing, i.e., hydrogen bond distances and energies ΔE_{int} (eqs 4 and 5) differ by only 0.01 Å and 0.2 kcal/mol, respectively. The same holds for methylation at the positions where the glucosidic N−C bond occurs in nucleosides (N9 in A en G, N1 in T and C): with bases methylated in this way, AT (**1c**) and GC(**2b**) hydrogen bond distances and energies differ only by up to 0.01 Å and 0.3 kcal/mol, respectively, from **1a** and **2a**.

Similarly, only very small effects occur on substituting hydrogen by 2′-deoxyribose (**1d** and **2c**) or neutral 2′-deoxyribose-5′-phosphate residues (**1e** and **2d**) at the N9 and N1 atoms of the purine and pyrimidine bases, respectively. For AT, the hydrogen bond energy decreases by only 0.3 kcal/mol on going from the Watson−Crick pair of plain nucleic bases (**1a**, −13.0 kcal/mol) to those of either nucleosides or nucleotides (**1d** and **1e**, both −12.7 kcal/mol; Table 5). At the same time, the N6−O4 and N1−N3 hydrogen bond distances go from 2.85 and 2.81 Å (**1a**) to 2.87 and 2.77 Å in **1d** and to 2.83 and 2.76 Å in **1e** (Figure 5). This does not resolve the discrepancy with the experimental values of 2.95 and 2.82 Å (**1**). And again, for GC, the hydrogen bond energy changes only slightly as we go from the Watson−Crick pair of plain nucleic bases (**2a**, −26.1 kcal/mol) to those of either nucleosides or nucleotides (**2c** and **2d**, both −25.3 kcal/mol; Table 6). The O6−N4, N1−N3 and N2−O2 hydrogen bond distances change only by up to 0.03 Å along **2a**, **2c** and **2d** (Figure 6). Thus, we still have the erroneous bond-length pattern of short−long−long at variance with the experimental order of long−long−short (**2**).

We went even one step further by studying the Watson−Crick complex of a strand of two nucleotides, namely that of deoxyadenylyl-3′,5′-deoxyuridine, i.e., $(dApdU)_2$ (**1k**). This is a model for the corresponding adenylyl-3′,5′-uridine complex $(ApU)_2$ in the crystal (**1**) studied by Seeman et al.[8a] (we have only removed the 2′-OH groups of ribose to somewhat reduce the immense computational cost). The structure of both our model $(dApdU)_2$ (**1k**) and the $(ApU)_2$ complex (**1**) is illustrated by Scheme 2.

The BP86/DZP geometry of **1k** is shown from different perspectives in Figure 7, left. As can be seen, the AU hydrogen bond distances in **1k** (Figure 7) differ only slightly, i.e., at most

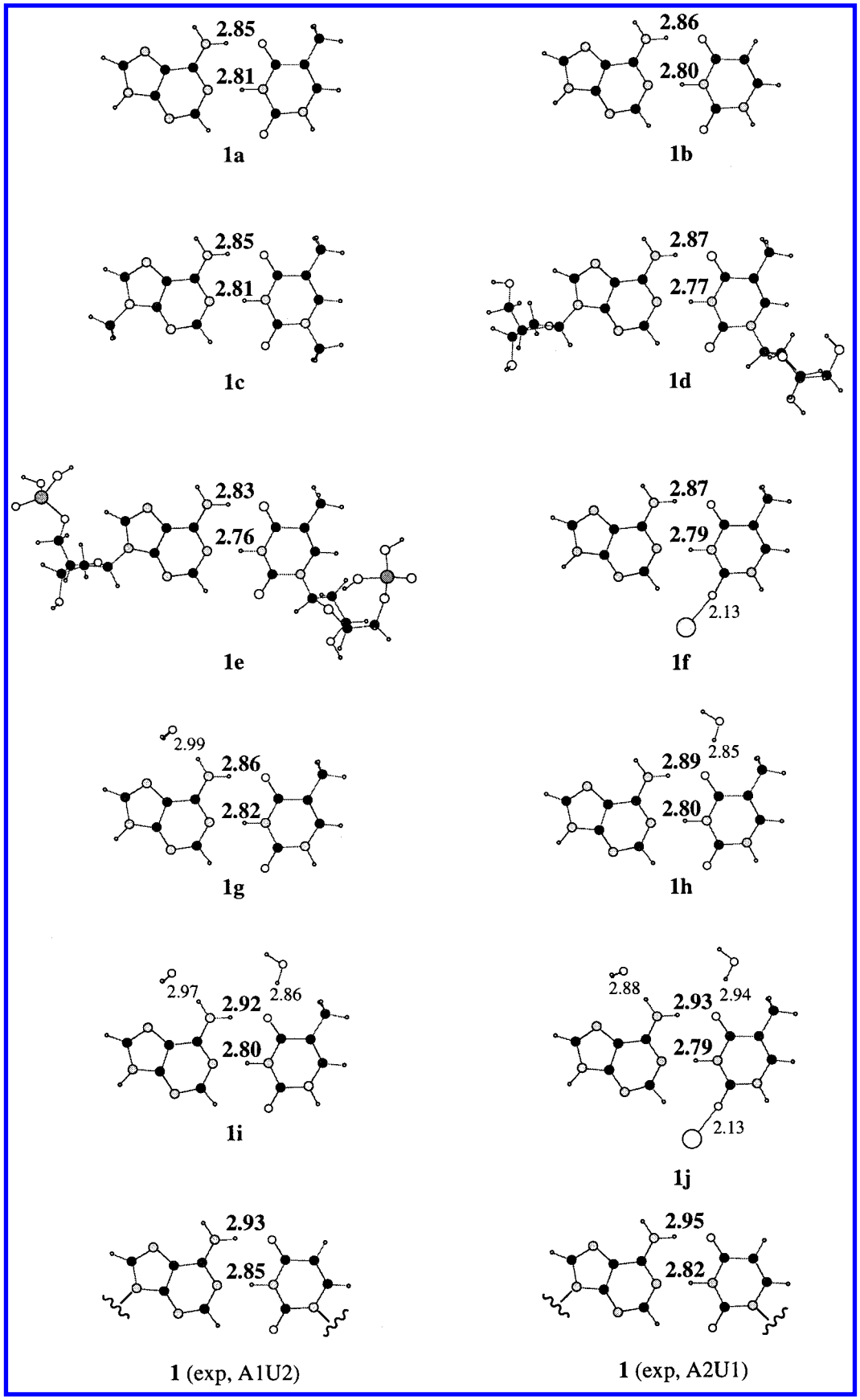


Figure 5. N6−O4 and N1−N3 distances in AT (**1a**), AU (**1b**), methylated AT (**1c**), AT with deoxyribose residues (**1d**), AT with deoxyribose 5′-phosphate residues (**1e**) and various AT crystal model systems (**1f**−**j**) from BP86/TZ2P (**1a**−**d**, **1f**−**j**) and BP86/DZP (**1e**) computations, and from the X-ray crystal structure of sodium adenylyl-3′,5′-uridine hexahydrate (**1**).[8a] Geometries of **1d**,**e** were optimized without any symmetry constraint whereas for the other systems (**1a**−**c** and **1f**−**i**) C_s symmetry has been used. We also show the distances between the oxygen of water and the proton-donor or proton-acceptor atom of the bases, and those between Na^+ and O2 of thymine. Note that there are two experimental values for both N6−O4 and N1−N3 because the two AU pairs in the crystal of **1** (see Scheme 2) experience different environments.

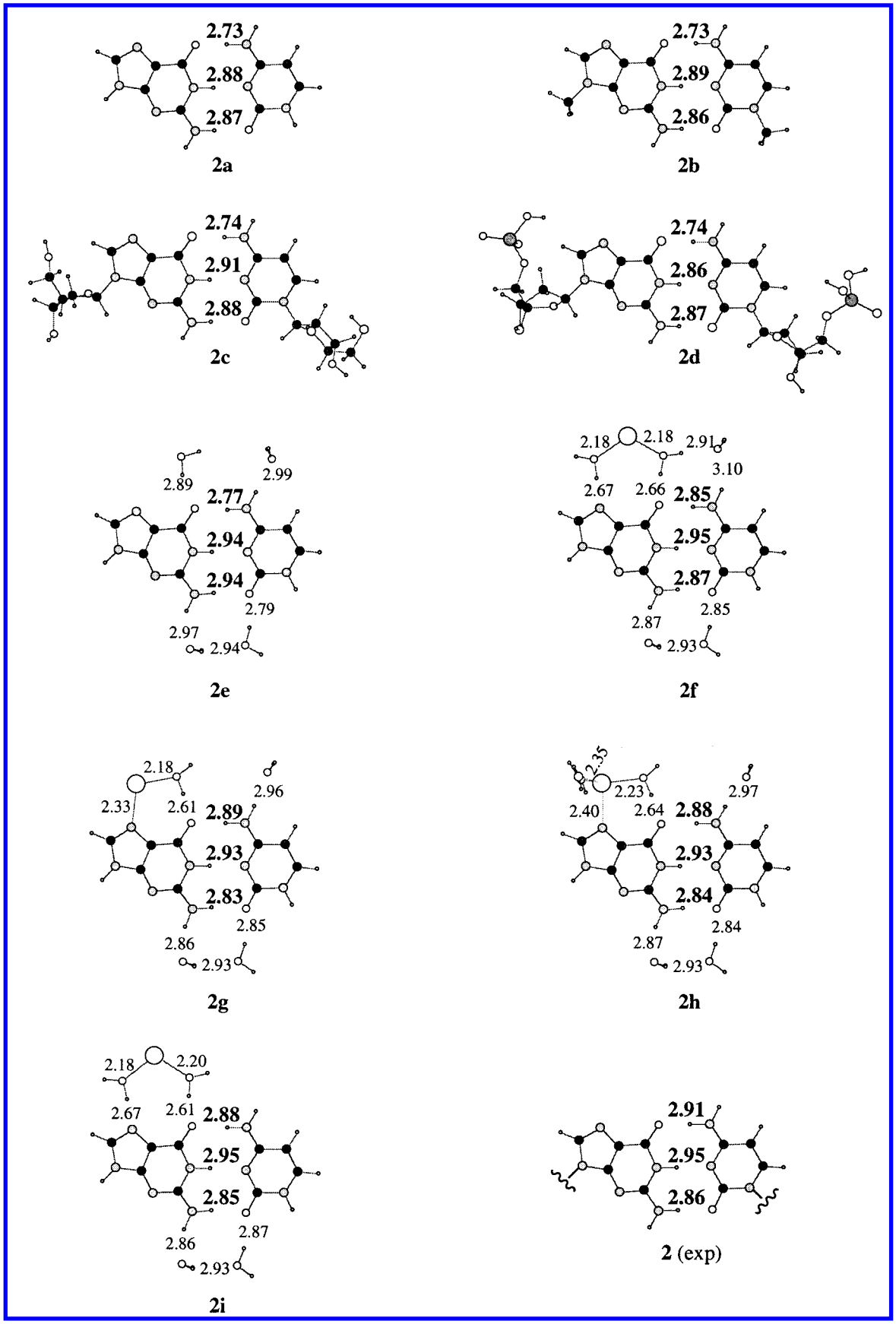


Figure 6. O6−N4, N1−N3, and N2−O2 distances in GC (**2a**), methylated GC (**2b**), GC with deoxyribose residues (**2c**), GC with deoxyribose 5′-phosphate residues (**2d**) and various GC crystal model systems (**2e**−**i**) from BP86/TZ2P (**2a**−**c**, **2e**−**i**) and BP86/DZP (**2d**) computations, and from the X-ray crystal structure of sodium guanylyl-3′,5′-cytidine nonahydrate (**2**).[8b] Geometries of **2c**,**d** were optimized without any symmetry constraint whereas for the other systems (**2a**,**b** and **2e**−**i**) C_s symmetry has been used. We also show the distances between the oxygen of water and the proton-donor or proton-acceptor atom of the bases, those between Na^+ and lone-pair donating atoms of guanine or water molecules, and those between oxygen atoms of water molecules.

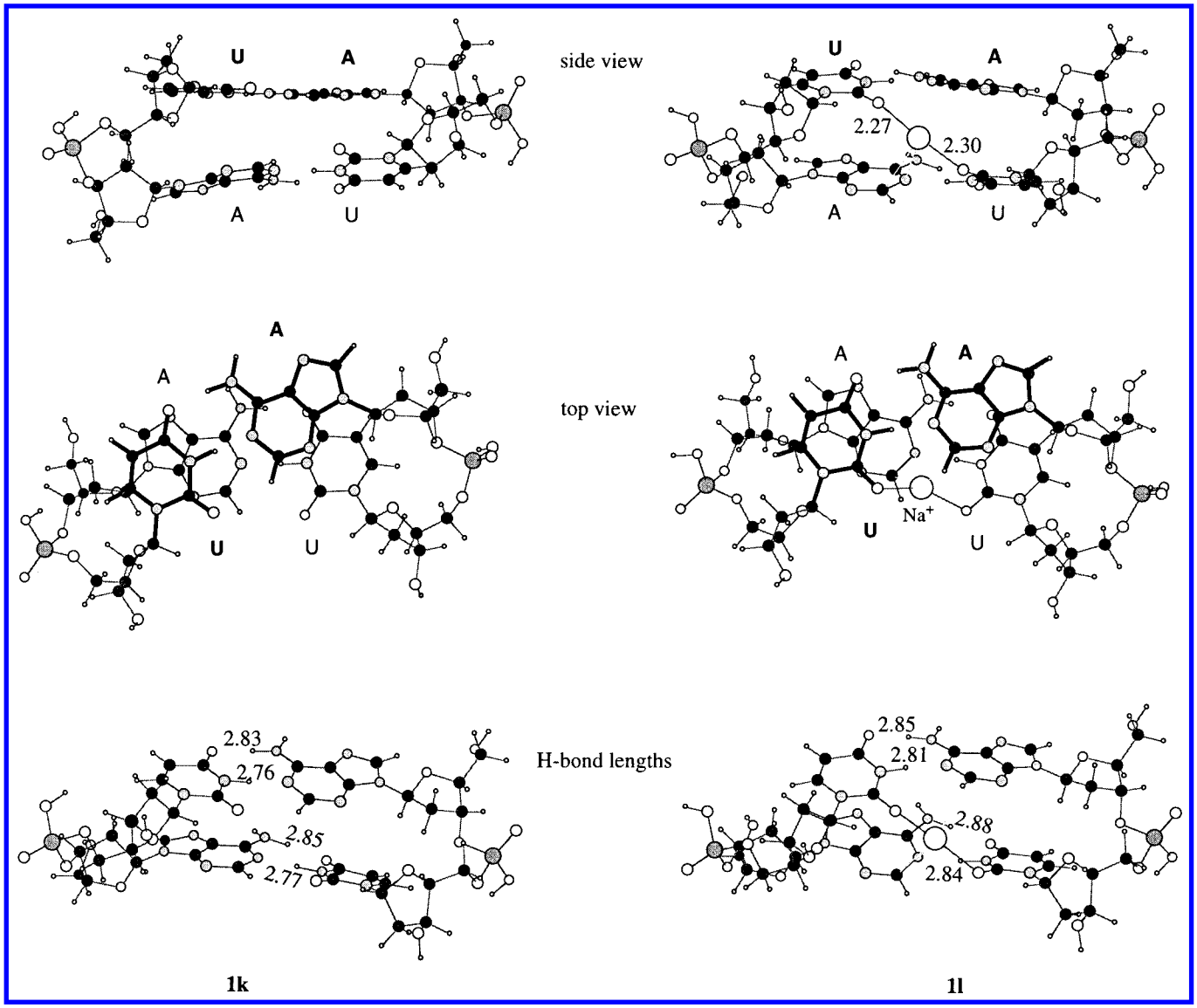


Figure 7. Different perspectives of the BP86/DZP structure of the Watson−Crick-type dimer of deoxyadenylyl-3′,5′-deoxyuridine, $(dApdU)_2$ (see also Scheme 2), with Na^+ ion (**1l**) and without Na^+ ion (**1k**), both optimized in C_1 symmetry without any symmetry constraint. The illustration shows N6−O4 and N1−N3 distances in AU pairs and the distances between the O2 atoms of each uracil base and the Na^+ ion.

Table 5. Analysis of the A-T Interaction (kcal/mol) in **1a**−**1k** (with Environment Effects in **1f**−**1j**)[a]

	1a	1b	1c	1d	1e[b]	1f	1g	1h	1i	1j	1k[b]
ΔE_{Pauli}	38.7	38.7	38.3	41.1	45.0	39.9	37.5	37.5	37.0	38.5	89.7
ΔV_{elstat}	−31.8	−32.0	−31.7	−32.9	−35.1	−33.4	−30.7	−30.7	−30.2	−33.2	−69.1
$\Delta E_{Pauli}+\Delta V_{elstat}$	6.9	6.7	6.6	8.2	9.9	6.5	6.8	6.9	6.8	5.3	20.6
ΔE_{σ}	−20.4	−20.5	−20.1			−24.3	−19.5	−19.8	−19.2	−23.2	
ΔE_{π}	−1.7	−1.7	−1.7			−4.2	−1.6	−1.6	−1.5	−4.0	
ΔE_{oi}	−22.1	−22.2	−21.8	−23.3	−25.9	−28.5	−21.1	−21.4	−20.7	−27.2	−50.7
ΔE_{int}	−15.2	−15.5	−15.2	−15.1	−16.0	−22.0	−14.3	−14.5	−13.9	−21.9	−30.1
ΔE_{prep}	2.2	2.3	2.1	2.4	3.3						9.2
ΔE	−13.0	−13.2	−13.1	−12.7	−12.7						−20.9

[a] BP86/TZ2P. See Figures 5 and 7. All bond energies relative to bases fully optimized in C_1 symmetry. [b] BP86/TZ2P//BP86/DZP.

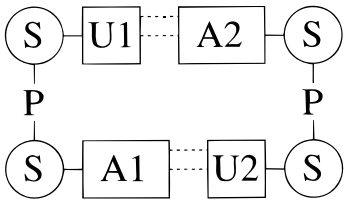
by 0.03 Å from those of plain AT (**1a**) also obtained at BP86/DZP (Table 3). The Watson−Crick-pairing energy ΔE of **1k** equals −20.9 kcal/mol at BP86/TZ2P// BP86/DZP (Table 5). Note that, although **1k** involves two AU pairs, this is significantly *less* than twice the pairing energy ΔE of AT (**1a**) or AU (**1b**). This can be ascribed to the strain in the backbone, which shows up in the much higher preparation energy ΔE_{prep} of 9.2 kcal/mol, and not to the actual interaction energy ΔE_{int} of −30.2 kcal/mol between the strands which, in fact, *is twice as strong* as that of a single base pair (Table 5).

Table 6. Analysis of the G-C Interaction (kcal/mol) in **2a**−**2d** (with Environment Effects in **2e**−**2i**)[a]

	2a	2b	2c	2d[b]	2e	2f	2g	2h	2i
ΔE_{Pauli}	51.9	51.1	48.6	53.7	47.7	51.9	44.7	45.0	43.2
ΔV_{elstat}	−48.5	−47.8	−46.0	−48.6	−50.7	−56.0	−51.6	−51.4	−46.5
$\Delta E_{\text{Pauli}}+\Delta V_{\text{elstat}}$	3.4	3.3	2.6	5.1	−3.0	−4.1	−6.9	−6.4	−3.3
ΔE_{σ}	−29.2	−28.9			−25.1	−28.1	−24.5	−24.5	−23.5
ΔE_{π}	−4.8	−4.6			−4.8	−4.1	−4.6	−4.6	−4.2
ΔE_{oi}	−34.0	−33.5	−32.1	−35.2	−29.9	−32.2	−29.1	−29.1	−27.7
ΔE_{int}	−30.6	30.2	−29.5	−30.1	−32.9	−36.3	−36.0	−35.5	−31.0
ΔE_{prep}	4.5	4.4	4.2	4.8					
ΔE	−26.1	−25.8	−25.3	−25.3					

[a] BP86/TZ2P. See Figure 6. All bond energies relative to bases fully optimized in C_1 symmetry. [b] BP86/TZ2P//BP86/DZP.

Scheme 2



In conclusion, the backbone has only a marginal influence on Watson−Crick hydrogen bonds and is thus not the source for the disagreement between theoretical and experimental structures mentioned above. But we can make use of this finding for reducing the computational cost in our further investigations on the effect of the crystal environment that the bases experience in **1** and **2** by leaving out the backbone.

5. The Effect of the Crystal Environment

5A. Environment Effects on Watson−Crick Structures. As will appear in the following, reconciliation of theory and experiment regarding AT and GC structures is achieved if one incorporates the effects of the molecular environment on the Watson−Crick pairs in the crystals of sodium adenylyl-3′,5′-uridine hexahydrate (**1**) and sodium guanylyl-3′,5′-cytidine nonahydrate (**2**) into the theoretical model systems. We begin with AT or AU. In **1**,[8a] the amino group of adenine and the O4 atom of uridine interact with two water molecules (A1U2, see Scheme 2) or with two 3′-ribose-OH groups of another $(ApU)_2$ complex (A2U1). We have modeled these interactions at BP86/TZ2P by introducing successively two water molecules at the corresponding positions in AT (compare **1a** and **1g**−**i** in Figure 5). The N1−N3 bond is not much affected but the N6−O4 expands, only slightly for one H_2O (**1g** and **1h**) but significantly for two water molecules (**1i**). This leads to hydrogen bond lengths in close agreement with experiment (**1i**: N6−O4 and N1−N3 are 2.92 and 2.80 Å).

The effect of sodium counterions is modest. In the crystal (**1**), one of the sodium ions bridges the O2 atoms of the two uracil bases (see Scheme 2 and Figure 7). First, we have modeled this by adding an Na^+ ion to plain (**1a**) and dihydrated AT (**1i**): the changes in hydrogen bond length in the resulting systems **1f** and **1j**, respectively, are marginal, i.e., 0.01−0.02 Å (Figure 5). The bridging Na^+ ion in the real crystal (**1**) may have a somewhat more pronounced effect because, there, it binds simultaneously to two uridine bases that are part of opposite strands of the double-helical segment and involved in different AU pairs (Figure 7). We have studied this at BP86/DZP by adding an Na^+ ion to $(dApdU)_2$ (**1l**, see Figure 7). Indeed, with hydrogen bond elongations of 0.02−0.07 Å on going from **1k** to **1l**, the effect is a bit more pronounced than in the case of the flat model systems (**1a**, **1f**, and **1j**). But eventually, the hydrogen bond lengths in **1l** still deviate significantly from the experimental values for **1** (compare Figures 5 and 7). We note that at variance with the situation in our model **1l**, the sodium ion in **1** does not enter into the space between the layers of the two AU pairs. Instead, it remains in the minor groove where it can bind also to water molecules.

Next, we consider the environment effects on the structure of GC pairs in **2**.[8b] Here, the N7, O6, and N2 positions of guanine are involved in hydrogen bonds with water molecules that, in case of N7 and O6, coordinate to a Na^+ ion. The O2 position of cytosine also forms a hydrogen bond with a water molecule whereas N4 hydrogen binds to a 3′-ribose-OH group of a neighboring $(GpC)_2$ complex. We have modeled the interactions of GC with its environment at BP86/TZ2P by introducing up to six water molecules and one sodium cation (compare **2a** and **2e**−**i** in Figure 6). This time, the sodium ion appears to be crucial. Introducing four water molecules (one at each position, O6 and N2 of guanine, and N4 and O2 of cytosine) leads to a significant elongation of the O6−N4, N1−N3, and N2−O2 hydrogen bonds which are now 2.77, 2.94, and 2.94 Å (**2e**) but we still have the wrong bond length pattern short−long−long (**2e**) instead of long−long−short in the crystal (**2**, see Figure 6). The situation improves significantly if, in addition, a sodium cation is introduced. This has been done in **2f**−**i**: all these model systems (but **2f**) show the correct hydrogen bond length pattern (long−long−short) with O6−N4, N1−N3, and N2−O2 distances that, especially for **2i**, agree excellently (i.e., within 0.01−0.03 Å) with the X-ray data (**2**: 2.91, 2.95, and 2.86 Å).

5B. Environment Effects on Watson−Crick Bond Strength. To analyze how the Watson−Crick interaction energy ΔE_{int} (eqs 4 and 5) is affected by the environment, we divide our model systems (**1f**−**j** and **2e**−**i**) into two subsystems, each of which consists of one of the bases plus the environment molecules that are closest to that base (see Figures 5 and 6). For example, the sodium ion in **1f** and **1j** belongs to thymine. First, we examine the interaction in AT systems (**1a**, **1f**−**j**, Table 5). The introduction of water molecules has little effect. In **1g**−**i**, the Watson−Crick interaction energy ΔE_{int} decreases only slightly (at most by 1.6 kcal/mol) with respect to **1a**. The presence of the Na^+ ion in **1f** and **1j** causes stronger orbital interactions as a result of which the hydrogen bond interaction energy ΔE_{int}

Table 7. Analysis of the Interaction Energy (kcal/mol) and Charge Transfer (Electrons) between AT and the Environment in **1f**−**1j**[a]

AT with	**1f** Na^+	**1g** H_2O on A	**1h** H_2O on T	**1i** 2 H_2O	**1j** 2 H_2O + Na^+
Interaction Energy Decomposition[b]					
ΔE_{Pauli}	12.1	13.7	11.0	24.7	35.6
ΔV_{elstat}	−32.9	−13.3	−10.1	−23.5	−54.5
$\Delta E_{Pauli}+\Delta V_{elstat}$	−20.8	.4	.9	1.2	−18.9
ΔE_{σ}	−10.4	−5.6	−5.1	−10.7	−20.9
ΔE_{π}	−8.2	−.5	−.6	−.9	−8.9
ΔE_{oi}	−18.6	−6.1	−5.7	−11.6	−29.8
ΔE_{int}	−39.4	−5.7	−4.8	−10.4	−48.7
VDD Charge of AT[c]					
ΔQ_{AT}	+0.04	−0.02	+0.03	+0.01	+0.01

[a] BP86/TZ2P. See Figure 5. [b] See section 2B. [c] See section 2C.

increases by some 7 kcal/mol. In the case of GC (**2a**, **2e**−**i**, Table 6), hydration and the introduction of a sodium cation leads in all cases to a moderate increase of 0.4−5.7 kcal/mol of the hydrogen bond interaction ΔE_{int} with respect to **2a**. This is caused by a slight increase of the electrostatic attraction (and a reduction of Pauli repulsion).

We conclude that hydration and counterions combined have a clearly visible effect on the hydrogen bond structure and strength of Watson−Crick pairs. Although we are with our model systems, of course, still far removed from the real crystal, we have been able to incorporate the most important interactions with the crystal environment, and this has brought theoretical and experimental structures into agreement. It is interesting to note that the species we have studied may also be conceived as microsolvated base pairs and, in this respect, they are also simple models for Watson−Crick systems that are exposed to hydration and ions under physiological conditions.

5C. Analysis of Interaction with Environment. Finally, we take a short look at the interaction between the Watson−Crick pairs and the environment. We want to know the strength and the nature of these intermolecular forces. For that purpose, we divide our AT and GC models **1f**−**j** and **2e**−**i** again into two subsystems using, however, a partitioning that differs from the above one. This time, the first subsystem is the plain base pair and the second subsystem consists of the surrounding water molecules and/or sodium cation. The analyses of both the interaction (section 2B) and the associated charge transfer (section 2C) between these fragments in the various AT and GC systems are summarized in Tables 7 and 8.

If the environment consists of only water molecules that hydrogen bind to the base pairs (as in **1g**−**i** and **2e**), the interaction ΔE_{int} is between −4.8 through −18.0 kcal/mol which comes down to roughly −5 kcal/mol per H_2O. Orbital interactions are relatively important here. Although they are about half as strong as the electrostatic interaction, they still are crucial for achieving net binding. As revealed by the VDD analyses, the interaction between the base pairs and the water molecules is accompanied by charge transfer from or to the environment: in **1g** ($\Delta Q_{AT} = -0.02$ e) and **1h** ($\Delta Q_{AT} = +0.03$ e) the AT pair accepts and donates electrons, respectively, from the water molecule (the VDD charge of a base pair, ΔQ_{AT} or ΔQ_{GC}, is computed as the sum of the VDD charges of all atoms that belong to that pair). In **1i** and **2e**, donation of electrons to and acceptance of electrons from the water molecules of the environment occur simultaneously and almost cancel (in **1i**: $\Delta Q_{AT} = +0.01$ e, and in **2e**: $\Delta Q_{GC} = +0.02$ e).

Table 8. Analysis of the Interaction Energy (kcal/mol) and the Charge Transfer (Electrons) between GC and the Environment in **2e**−**2i**[a]

GC with	**2e** 4 H_2O	**2f** 5 H_2O + Na^+	**2g** 4 H_2O + Na^+	**2h** 6 H_2O + Na^+	**2i** 4 H_2O + Na^+
Interaction Energy Decomposition[b]					
ΔE_{Pauli}	32.8	65.9	61.2	57.6	65.0
ΔV_{elstat}	−32.2	−68.0	−84.2	−75.1	−67.9
$\Delta E_{Pauli}+\Delta V_{elstat}$	.6	−2.1	−23.0	−17.5	−2.9
ΔE_{σ}	−17.1	−42.0	−35.8	−31.9	−43.0
ΔE_{π}	−1.5	−6.1	−8.2	−6.2	−6.9
ΔE_{oi}	−18.6	−48.1	−44.0	−38.1	−49.9
ΔE_{int}	−18.0	−50.2	−67.0	−55.6	−52.8
VDD Charge of GC[c]					
ΔQ_{GC}	+0.02	+0.15	+0.08	+0.07	+0.18

[a] BP86/TZ2P. See Figure 6. [b] See section 2B. [c] See section 2C.

Introducing a sodium cation into the environment (as in **1f**, **1j**, and **2f**−**i**) increases all components of the interaction energy (Tables 7 and 8). If Na^+ interacts directly with the base pair (**1f**, **1j**, **2g**, and **2h**), the electrostatic interaction gains in importance in the sense that it can overcome the Pauli repulsion and provide net bonding on its own. This is still substantially reinforced by sizable orbital interactions ΔE_{oi} (roughly half as strong as ΔV_{elstat}) that lead to a charge transfer of up to 0.18 electrons in the case of **2i** from base pair to environment. On the other hand, when the sodium cation is separated from the base pair by a shell of water molecules (as in **2f** and **2i**), the electrostatic interaction ΔV_{elstat} becomes smaller with respect to the other components of the interaction and merely compensates the Pauli repulsion. In these cases, both ΔV_{elstat} and ΔE_{oi} are needed to achieve substantial net bonding between base pair and environment.

6. Conclusions

We have unraveled a hitherto unresolved discrepancy between theoretical and experimental hydrogen bond lengths in Watson−Crick base pairs. The disagreement was caused by a deficiency in the model systems used so far in theoretical computations, namely, the absence of the molecular environment (i.e., water, sugar OH groups, counterions) that the base pairs experience in the crystals studied experimentally. If we incorporate the major elements of this environment into our model, simulating them by up to six water molecules and one Na^+ ion, we achieve excellent agreement with experiment at BP86/TZ2P.

On the other hand, whether plain nucleic bases or more realistic models for the nucleotides are used is much less important. Neither hydrogen bond lengths nor strengths are significantly affected if we use methyl, ribose, or 5′-ribose monophosphate instead of hydrogen as the substituents at N9 and N1 of the purine and pyrimidine bases, respectively. Even the Watson−Crick-type dimer of deoxyadenylyl-3′,5′-deoxyuridine [$(dApdU)_2$, a model for a DNA segment of two base pairs] yields hydrogen bond lengths that differ only slightly from those in a plain AT or AU base pair.

Furthermore, we find that the BP86 functional yields A−T and G−C bond enthalpies in excellent agreement with experiment, especially in combination with the TZ2P basis. But also the smaller and, thus, more economic DZP basis leads to satisfactory results, provided that bond energies are corrected for the basis set superposition error. On the other hand, PW91

and BLYP functionals furnish hydrogen bonds that are up to 0.03 Å shorter and up to 2.5 kcal/mol more binding than those obtained with BP86.

Finally, our finding that present-day density functional theory is very well able to adequately describe biologically relevant molecules involving hydrogen bonds may have important consequences for future quantum biochemical studies. It indicates that DFT may become an efficient alternative to traditional (i.e., Hartree−Fock-based) ab initio methods for tackling this type of computationally extremely demanding problems. An important question that remains is how well in general DFT copes with systems that involve stacking between DNA base pairs.

Acknowledgment. F.M.B. thanks the Deutsche Forschungsgemeinschaft (DFG) for a Habilitation Fellowship and the Fonds der Chemischen Industrie (FCI) for financial support. The authors thank the National Computing Facilities (NCF) foundation of The Netherlands foundation for Scientific Research (NWO) for a grant of supercomputer time.

JA993262D